CLAIMS

What is claimed is:

1. A method for identifying desirable protein backbone configurations comprising:

generating backbone protein configurations using a set of dihedral angle pairs;

normalizing the total surface exposure of each remaining configuration;

generating a random set of sequences of hydrophobicities with uniform weight on the space of allowed sequences;

determining, for each randomly generated sequence, which of the remaining configurations is the ground state, and;

recording a ground-state configuration for each sequence wherein the desirable configurations are those containing the most sequences with that configuration as their ground state.

2. A method for identifying desirable protein backbone configurations as in claim 1 wherein:

one pair of dihedral angle pairs corresponds to an alpha helix and one pair of dihedral angle pairs corresponds to a beta strand.

3. A method for identifying desirable protein backbone configurations as in claim 1 wherein:

two sets of dihedral angles correspond to an alpha helix and one set of dihedral angle pairs corresponds to a beta strand.

4. A method for identifying desirable protein backbone configurations as in claim 3 wherein:

additional dihedral angles fall within regions of high frequency in a Ramachandran plot.

5. A method for identifying desirable protein backbone configurations as in claim 4 wherein:

the probability of choosing a particular pair of dihedral angles depends on the preceeding pairs of dihedral angles along the backbone.

6. A method for identifying desirable protein backbone configurations as in claim 5 further comprising:

eliminating self-intersecting configurations.

7. A method for identifying desirable protein backbone configurations as in claim 6 further comprising:

eliminating non-compact configurations.

8. A method for identifying desirable protein backbone configurations as in claim 7 further comprising:

clustering configurations sufficiently similar in the three dimensional trajectory followed by their backbones and treating all configurations within a cluster as variants of a single configuration, and;

summing, for all configurations in a cluster, the number of sequences with that configuration as their ground state such that the sum is considered the designability of the cluster.

9. A method for identifying desirable protein backbone configurations as in claim 8 further comprising:

eliminating configurations with low Variances.

10. A method for identifying desirable protein backbone configurations as in claim 1 wherein:

the set of dihedral angle pairs is a set of strings of dihedral angle pairs.

11. A method for identifying desirable protein backbone configurations as in claim 10 wherein:

the strings of angles are weighted according to their frequency of appearance in natural proteins and infrequent strings are eliminated.

12. A method for identifying desirable protein backbone configurations as in claim 1 wherein:

normalizing is accomplished by dividing the surface exposure of each amino acid in a given configuration by the total surface exposure of that configuration.

13. A method for identifying desirable protein backbone configurations as in claim 1 further comprising:

eliminating configurations with low Variance.

14. A method for identifying desirable protein backbone configurations as in claim 1 further comprising:

eliminating self-intersecting configurations.

15. A method for identifying desirable protein backbone configurations as in claim 14 further comprising:

eliminating non-compact configurations.

16. A method for identifying desirable protein backbone configurations as in claim 1 further comprising:

eliminating non-compact configurations.

17. A method for identifying desirable protein backbone configurations as in claim 1 further comprising:

eliminating configurations with low Variance.

18. A method for identifying desirable protein backbone configurations as in claim 1 further comprising:

eliminating all configurations that are not favorable for forming a large number of hydrogen bonds after eliminating non-compact configurations.

19. A method for identifying desirable protein backbone configurations as in claim 1 further comprising:

clustering configurations sufficiently similar in the three dimensional trajectory followed by their backbones and treating all configurations within a cluster as variants of a single configuration, and;

summing, for all configurations in a cluster, the number of sequences with that configuration as their ground state such that the sum is considered the designability of the cluster.

20. A method for identifying desirable protein backbone configurations as in claim 19 wherein:

clustering is accomplished by totaling the root-mean-square distance between every pair of configurations and defining a configuration as a member of a cluster if it lies within a root-mean-square distance λ of any member of the cluster.

21. A method for identifying desirable protein backbone configurations as in claim 20 wherein:

λ is 0.4 Å per amino acid.

22. A method for designing proteins comprising:

generating backbone protein configurations using a set of dihedral angle pairs;

eliminating self-intersecting configurations;

normalizing the total surface exposure of each remaining configuration;

generating a random set of sequences of hydrophobicities with uniform weight on the space of allowed sequences for each remaining configuration;

determining, for each randomly generated sequence, which of the remaining configurations is the ground state;

recording the ground-state configuration for each sequence wherein desirable configurations are those containing the most sequences with that configuration as their ground state, and;

synthesizing sequences of amino acids for the desirable configurations.

23. A method for designing proteins as in claim 22 wherein:

one set of dihedral angle pairs corresponds to an alpha helix and one set of dihedral angles corresponds to a beta strand.

24. A method for designing proteins as in claim 22 wherein:

two sets of dihedral angles correspond to an alpha helix and one set of dihedral angle pairs corresponds to a beta strand.

25. A method for designing proteins as in claim 24 wherein:

additional dihedral angle pairs fall within regions of high frequency in a Ramachandran plot.

26. A method for designing proteins as in claim 25 wherein:

the probability of choosing a particular pair of dihedral angles depends on the preceeding pairs of dihedral angles along the backbone.

27. A method for designing proteins as in claim 26 further comprising:

eliminating self-intersecting configurations.

28. A method for designing proteins as in claim 27 further comprising:

eliminating non-compact configurations.

29. A method for designing proteins as in claim 28 further comprising:

clustering configurations sufficiently similar in the three dimensional trajectory followed by their backbones and treating all configurations within a cluster as variants of a single configuration, and;

summing, for all configurations in a cluster, the number of sequences with that configuration as their ground state such that the sum is considered the designability of the cluster.

30. A method for designing proteins as in claim 29 further comprising:

recording the Variance of each configuration, ranking the configurations from highest Variance to lowest, and

designing proteins starting with the configurations having the highest Variance.

31. A method for designing proteins as in claim 22 wherein:

the set of dihedral angles is a set of strings of dihedral angles.

32. A method for designing proteins as in claim 31 wherein:

the strings of angles are weighted according to their frequency of appearance in natural proteins and infrequent strings are eliminated.

33. A method for designing proteins as in claim 22 wherein:

normalizing is accomplished by dividing the surface exposure of each amino acid in a given configuration by the total surface exposure of that configuration.

34. A method for designing proteins as in claim 22 further comprising:

recording the Variance of each configuration, ranking the configurations from highest Variance to lowest, and

designing proteins starting with the configurations having the highest Variance.

35. A method for designing proteins as in claim 22 further comprising:

eliminating non-compact configurations after self-intersecting configurations are eliminated.

36. A method for designing proteins as in claim 35 further comprising:

clustering configurations sufficiently similar in the three dimensional trajectory followed by their backbones and treating all configurations within a cluster as variants of a single configuration, and;

summing, for all configurations in a cluster, the number of sequences with that configuration as their ground state such that the sum is considered the designability of the cluster.

37. A method for designing proteins as in claim 22 further comprising:

eliminating all configurations that are not favorable for forming hydrogen bonds after eliminating non-compact configurations.

38. A method for designing proteins as in claim 22 further comprising:

clustering configurations sufficiently similar in the three dimensional trajectory followed by their backbones and treating all configurations within a cluster as variants of a single configuration, and;

summing, for all configurations in a cluster, the number of sequences with that configuration as their ground state such that the sum is considered the designability of the cluster.

39. A method for designing proteins as in claim 38 wherein:

clustering is accomplished by totaling the root-mean-square distance between every pair of configurations and defining a configuration as a member of a cluster if it lies within a root-mean-square distance λ of any member of the cluster.

40. A method for designing proteins as in claim 39 wherein:

λ is 0.4 Angstroms per amino acid.

41. A method for analyzing the designability of protein backbone configurations to determine if the number of sequences each configuration has in its ground state is larger than a predetermined number comprising:

generating backbone protein configurations using a set of dihedral angle pairs;

eliminating self-intersecting configurations;

normalizing the total surface exposure of each remaining configuration;

generating a random set of sequences of hydrophobicities with uniform weight on the space of allowed sequences;

determining, for each randomly generated sequence, which of the remaining configurations is the ground state;

recording a ground-state configuration for each sequence wherein the desirable configurations are those containing the most sequences with that configuration as their ground state, and;

comparing how many sequences each configuration has in its ground-state with the predetermined number whereby configurations with larger numbers are highly designable.

42. A method for analyzing the designability of protein backbone configurations as in claim 41 wherein:

normalizing is accomplished by dividing the surface exposure of each amino acid in a given configuration by the total surface exposure of that configuration.

43. A method for analyzing the designability of protein backbone configurations as in claim 41 wherein:

one set of dihedral angle pairs corresponds to an alpha helix and one set of dihedral angle pairs corresponds to a beta strand.

44. A method for analyzing the designability of protein backbone configurations as in claim 41 wherein:

two sets of dihedral angle pairs correspond to an alpha helix and one set of dihedral angle pairs corresponds to a beta strand.

45. A method for analyzing the designability of protein backbone configurations as in claim 44 wherein:

additional dihedral angles fall within regions of high frequency in a Ramachandran plot.

46. A method for analyzing the designability of protein backbone configurations as in claim 45 wherein:

the probability of choosing a particular pair of dihedral angles depends on the preceeding pairs of dihedral angles along the backbone.

47. A method for analyzing the designability of protein backbone configurations as in claim 46 further comprising:

eliminating non-compact configurations after self-intersecting configurations are eliminated.

48. A method for analyzing the designability of protein backbone configurations as in claim 47 further comprising:

recording the Variance of each configuration, ranking the configurations from highest Variance to lowest, and

designing proteins starting with the configurations having the highest Variance.

49. A method for analyzing the designability of protein backbone configurations as in claim 48 further comprising:

clustering configurations sufficiently similar in the three dimensional trajectory followed by their backbones and treating all configurations within a cluster as variants of a single configuration, and;

summing, for all configurations in a cluster, the number of sequences with that configuration as their ground state such that the sum is considered the designability of the cluster.

50. A method for analyzing the designability of protein backbone configurations as in claim 41 wherein:

the set of dihedral angle pairs is a set of strings of dihedral angle pairs.

51. A method for analyzing the designability of protein backbone configurations as in claim 49 wherein:

the strings of angles are weighted according to their frequency of appearance in natural proteins and infrequent strings are eliminated.

52. A method for analyzing the designability of protein backbone configurations as in claim 41 wherein:

the probability of choosing a particular pair of dihedral angles depends on the preceeding pairs of dihedral angles along the backbone.

53. A method for analyzing the designability of protein backbone configurations as in claim 41 further comprising:

recording the Variance of each configuration, ranking the configurations from highest Variance to lowest, and

designing proteins starting with the configurations having the highest Variance.

54. A method for analyzing the designability of protein backbone configurations as in claim 41 further comprising:

eliminating all configurations that are not favorable for forming a large number of hydrogen bonds after eliminating non-compact configurations.

55. A method for analyzing the designability of protein backbone configurations as in claim 41 further comprising:

clustering configurations sufficiently similar in the three dimensional trajectory followed by their backbones and treating all configurations within a cluster as variants of a single configuration, and;

summing, for all configurations in a cluster, the number of sequences with that configuration as their ground state such that the sum is considered the designability of the cluster.

56. A method for analyzing the designability of protein backbone configurations as in claim 55 further comprising:

clustering is accomplished by totaling the root-mean-square distance between every pair of configurations and defining a configuration as a member of a cluster if it lies within a root-mean-square distance λ of any member of the cluster.

57. A method for analyzing the designability of protein backbone configurations as in claim 56 further comprising:

λ is 0.4 Angstroms per amino acid.